Applied Bayesian Statistics

Teachers: Herbert Hoijtink, Ellen Hamaker, Milica Miočević, Caspar van Lissa, and Lion Behrens
Department of Methods and Statistics
P.O.Box 80140
3508 TC Utrecht

Abstract

The interest in Bayesian statistic has been increasing for the last two decades both among statisticians and researchers who cannot properly analyze their data using methods based on classical statistics. This course will provide a sound basis in Bayesian statistics for those who

- Want to understand what Bayesian statistics is about
- Want to use Bayesian statistics to build and evaluate statistical models
- Want to get hands on experience with Bayesian statistics in OpenBUGS or JAGS and Bain

Target Group

The course is aimed at researchers who not only work with statistical tools, but are also interested in the development and evaluation of statistical tools. Among these are psychometricians, sociometricians, epidemiologists, and statisticians. The only requirement is familiarity with the following concepts: the likelihood function, the p-value, analysis of variance, and multiple regression.

Course Data, Course Fee and Course Materials

The course will be given April 30 – May 4, 2018 in Utrecht at the Uithof. Each day will start at 9.00 and finish at 17.00.

Course materials will be made available one week in advance. The reading materials listed at the end of this document are freely available via the library of most universities and can be downloaded for further reading. The books listed in the reference list can also be considered further reading, and can be obtained in most academic book shops.

Participants are required to bring their own lap-top on which either OpenBugs (http://www.openbugs.info/w/), or R (https://cran.r-project.org/) and JAGS (http://mcmc-jags.sourceforge.net/) should be installed. For mac users the latter option is recommended due to the limitations of openbugs. Participants should also install Bain by downloading the zip file from https://informative-hypotheses.sites.uu.nl/software/bain/ .

Grading

Students who want to obtain a grade and credit points (Epidemiology Master Students 1.5 EC, IOPS PhD students 2 EC) for this course have to attend the grading session on Friday. Each student can choose one of the following options:

- Present a theorem and defend it. This could, for example, be a position in the controversy between classical and Bayesian statistics.
- Present an exercise executed during this course and elaborate it using a personalized approach.
- Present a Bayesian analysis performed on your own data.
- Other ideas are appreciated. However, verify with the instructor before you start working on it.

Day 1: Introduction to the Key Concepts and Formulas of Bayesian Statistics: Introduction to OpenBUGS.

Using simple models for binomial data a technical introduction of the key concepts in Bayesian statistics will be given. This day consists of three lectures and a lab meeting in which the students get acquainted with OpenBUGS and practice what they have learned during the lectures.

Lecture 1: Using an example in which the data have a binomial distribution, key concepts in Bayesian statistics will be introduced: prior distribution, density of the data, posterior distribution, expected a posteriori estimates, credibility intervals and sampling from the posterior distribution.

Lecture 2: The posterior predictive distribution will be discussed. This distribution lies at the heart of the Bayesian equivalent of classical hypothesis testing: posterior predictive p-values. The pros and cons of posterior predictive compared to classical p-values will be discussed.

Lab Meeting: The students will learn to work with OpenBUGS. This will be done by means of an application of the key concepts and formulas of Bayesian statistics on data that have a binomial distribution. This lab meeting will also be used to ensure that OpenBUGS works properly on your laptop.

Day 2: Bayesian Estimation: the Gibbs sampler and the Metropolis-Hastings Algorithm. Bayesian Model Selection: the DIC.

On Day 1 it was shown that it is easy to sample from posterior distributions that are based on simple binomial densities for the data. However, for more complicated models/densities direct sampling is often not possible. In two lectures and one lab meeting it will be explained how this problem can be solved. Furthermore, an alternative for the Bayes factor will be discussed in the third lecture: the DIC.

Lecture 1: The Gibbs sampler: sampling from a multivariate posterior distribution if the conditional densities are known. The Metropolis Hastings algorithm: sampling from a multivariate posterior distribution if the conditional densities are unknown. Convergence issues when Markov chain Monte Carlo methods are used to sample from a posterior distribution.

Lecture 2: In this lecture the ideas underlying the deviance information criterion DIC will be discussed. Illustration of applications of the DIC will be provided.

Lab Meeting: OpenBUGS will be used to construct a Gibbs sampler for the posterior distribution of a logistic regression model. The students will practice using the DIC in the context of the selection of predictors for the logistic regression model. Furthermore, for a person with known predictors but unknown outcome, a prediction of and credibility interval for the unknown outcome will be obtained using OpenBUGS.

Day 3: Bayesian Model Selection: the Bayes Factor.

Today the Bayesian alternative for hypothesis testing will be discussed: Bayesian model selection. There will be lectures and a lab meeting.

Lecture 1: Informative hypotheses and the Bayes factor will be introduced. All from an applied perspective.

Lecture 2: Bayesian error probabilities, Bayesian updating, and the evaluation of replication studies using informative hypotheses will be discussed. All from an applied perspective.

Lecture 3: A technical elaboration of the Bayes factor in the context of a multiple regression model.

Lab Meeting: Using Bain the participants will formulate and evaluate informative hypotheses in the context of an ANOVA model.

Day 4: Bayesian linear regression analysis: evaluating moderation and mediation effects.

Today's material will focus on Bayesian methods for parameter estimation in linear models. There will be three lectures and a lab meeting.

Lecture 1: Bayesian linear regression analysis: model specification, prior selection, estimation, and interpretation of effects. This lecture is a basis for the subsequent two lectures on moderation and mediation analysis in the Bayesian framework.

Lecture 2: Bayesian moderation analysis will be introduced: model specification, prior selection, estimation, and interpretation of effects.

Lecture 3: Bayes mediation analysis will be introduced: model specification, prior selection, estimation, and interpretation of effects.

Lab meeting: Using OpenBUGS/JAGS the participants will learn how to fit linear models with moderation and mediation effects in the Bayesian framework.

Day 5: Some philosophical background of Bayesian and Frequentist statistics

In today's morning session we will zoom in more on the philosophical background of Bayesian and frequentist statistics. For this purpose we will have an interactive lecture, where we exchange thoughts about the role of Frequentist and Bayesian statistics in the scientific method, and we will host a debate where the students will be divided into Frequentist and Bayesian groups and have to promote their respective convictions. Students may find inspiration for this debate in the assigned reading for Friday.

Grading sessions will take place in the afternoon.

Reading

Contained in the Reader

Hoijtink, H. (2009). Bayesian Data Analysis. In: R.E. Millsap and A. Maydeu-Olivares, The SAGE Handbook of Quantitative Methods in Psychology. London: SAGE.

Klugkist, I., Laudy, O. and Hoijtink, H. (2005). Inequality Constrained Analysis of Variance: A Bayesian approach. *Psychological Methods, 4,* 477-493.

Recommended Reading for Day 1

Gelman, A., Meng, X. and Stern, H.S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica, 6*, 733-760.

Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. foundations and basic methods. *International Journal of Epidemiology*, 35(3):765-775.

Rietbergen, C., Klugkist, I., Janssen, K.J.M., Hoijtink, H., and Moons, K.G.M. (2011). Incorporation of Historical Data in the Analysis of Randomized Therapeutic Trials. *Contemporary Clinical Trials, 6,* 848-855

Recommended Reading for Day 2

Hamaker, E. L., Van Hattum, P., Kuiper, R. M. & Hoijtink, H. (in press). Model selection based on information criteria in multilevel modeling. In K. Roberts & J. Hox (Eds). *Handbook of Advanced Multilevel Analysis.* Taylor and Francis.

Greenland, S. (2007). Bayesian perspectives for epidemiological research. II. regression analysis. *International Journal of Epidemiology*, 36(1):195-202.

An application of Bayesian Statistics

Cerdá, M., Tracy, M, Messner, S.F., Vlahov, D., Tardiff, K., and Galea, S. (2009). Misdemeanor policing, physical disorder, and gun-related homicide: A spatial analytic test of "Broken-Windows" theory. *Epidemiology*, 20: 533–541

Recommended Reading for Day 3

The tutorial (and the references contained in it) that can be downloaded from
https://informative-hypotheses.sites.uu.nl/software/bain/
You will find the tutorial in the .zip file containing the most recent version of Bain.

An application of informative hypotheses

Van Well, S., Kolk, A.M. and Klugkist, I.G. (2008). Effects of Sex, Gender Role Identification, and Gender relevance of Two Types of Stressors on Cardiovascular and Subjective Responses: Sex and Gender Match/Mismatch Effects. *Behavior Modification, 32,* 427-449.

Recommended Reading for Day 4

Wang, L., & Preacher, K. J. (2015). Moderated mediation analysis using Bayesian methods. Structural Equation Modeling: A Multidisciplinary Journal, 22(2), 249-263.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. Psychological methods, 14(4), 301-322.

Recommended Reading for Day 5

Berger, J., & Berry, D. (1988). "Statistical Analysis and the Illusion of Objectivity." *American Scientist*, 76: 159-165.

Cohen, J. (1994). "The Earth Is Round (p<.05)." *American Psychologist*, 49 (12), 997-1003.

Dennis, B. (1996). "Discussion: Should Ecologists Become Bayesians?" *Ecological Applications*, 6 (4), 1095-1103.

Gelman, A. (2008). "Objections to Bayesian Statistics." *Bayesian Analysis*, 3 (3), 445-450.